Taylor & Francis
Taylor & Francis Group

Journal of Quality Technology
A Quarterly Journal of Methods, Applications and Related Topics

ISSN: 0022-4065 (Print) 2575-6230 (Online) Journal homepage: https://www.tandfonline.com/loi/ujqt20

A Comparison of Multivariate Control Charts for Individual Observations

Joe H. Sullivan & William H. Woodall

To cite this article: Joe H. Sullivan & William H. Woodall (1996) A Comparison of Multivariate Control Charts for Individual Observations, Journal of Quality Technology, 28:4, 398-408, DOI: 10.1080/00224065.1996.11979698

To link to this article: https://doi.org/10.1080/00224065.1996.11979698

Published online: 21 Feb 2018.

This paper was presented at the Journal of Quality Technology Session at the 40th Annual Fall Technical Conference of the Chemical and Process Industries Division and the Statistics Division of the American Society for Quality Control and the Section on Physical & Engineering Sciences of the American Statistical Association in Scottsdale, Arizona, October 24–25, 1996.

A Comparison of Multivariate Control Charts for Individual Observations

JOE H. SULLIVAN

Mississippi State University, Mississippi State, MS 39762

WILLIAM H. WOODALL

University of Alabama, Tuscaloosa, AL 35487-0226

For multivariate statistical process control with individual observations the usually recommended procedure in the retrospective phase is the Hotelling's T^2 control chart. All the observations are pooled to estimate the mean vector and covariance matrix. An out-of-control signal results if the T^2 value of any observation exceeds an upper control limit. We show this procedure is *not* effective in detecting a shift in the mean vector because the covariance matrix is badly estimated. Somewhat surprisingly, the signal probability decreases with increasingly severe shifts in the mean vector. We compare several alternative methods for estimating the covariance matrix and recommend a procedure analogous to the use of moving ranges in the univariate case. This procedure uses the vector differences between successive observations to estimate the in-control covariance matrix of the process. With this more robust estimate, step and ramp shifts in the mean vector are more likely to be detected in the retrospective phase.

Introduction

USING the terminology of Alt and Smith (1988), multivariate statistical quality monitoring can be divided into the following parts:

Phase I: need to estimate process parameters
- Stage 1: *retrospective* examination of subgroup behavior
- Stage 2: *prospective* examination of *future* subgroups

Phase II: use of specified values for process parameters.

In the retrospective Stage 1 situation, historical observations are analyzed to decide if the process is in statistical control and to estimate the in-control parameters of the process. The difficulty is that the parameter estimates may be affected by special causes, possibly masking their presence. In the prospective Stage 2, control charts are used to detect departures from the statistical parameters estimated in Stage 1.

Sometimes the multivariate data can be grouped into rational subgroups, relying on properties of the production process that create homogeneity within the subgroups. When rational subgroups are present, a shift in the mean vector is presumed to be more likely to take place between subgroups than within a subgroup. This can be used to advantage by forming the sample covariance matrix for each subgroup, then averaging these to get an estimate of the process covariance matrix. Then the mean vectors for each subgroup can be examined for a shift, thereby detecting assignable causes for shifts in the mean vector.

However, sometimes the rational subgroup size is one, that is to say, the data are structured only as

---

Dr. Sullivan is Assistant Professor in Quantitative Analysis. He is a Member of ASQC.

Dr. Woodall is Russell Professor of Statistics in the Department of Management Science and Statistics. He is a Senior Member of ASQC.

individual observations, and process characteristics do not necessarily produce homogeneous subgroups of larger size. See Montgomery (1996, pp. 221–222) for several examples. In the case of individual observations, a seemingly reasonable approach is to chart Hotelling's T^2 statistic, using the sample mean vector and covariance matrix. If any values exceed an upper control limit, an out-of-control signal is generated. This method is recommended by authors who treat the case of individual multivariate observations. For example, see Jackson (1985); Tracy, Young, and Mason (1992); Lowry and Montgomery (1995); or Wierda (1994a). Surprisingly, as we show later, this procedure is insensitive to a step or ramp change in the mean vector.

In the univariate case it has long been recognized that the variance estimated from moving ranges is more robust than the sample variance in the presence of a shift in the mean. (See, for example, Nelson (1982) or Montgomery (1996, p. 222).) However, an important source of this robustness is the division of the data into successive pairs that tend to have the same or nearly the same mean. For two observations the sample standard deviation is the range divided by the square root of two, so using "moving standard deviations" in the univariate case would be equally robust.

We compare control charts using several alternative methods to estimate the covariance matrix and recommend a more powerful procedure that takes advantage of the serial nature of the observations. With a step or ramp change in the mean vector, adjacent observations are likely to have the same or nearly the same mean vector. With a step change, only one pair of successive observations will have different mean vectors, while all the other pairs will have the same mean vector. With a ramp change, successive pairs will have a difference in mean vectors that is small compared to that of randomly paired observations. To use this structure to advantage, one can simply take the vector differences between successive observations. One-half times the sample covariance matrix of these differences is used to estimate the process covariance matrix. We provide theoretical and performance-based justification for this procedure that was initially suggested by Holmes and Mergen (1993) and also described by Wierda (1994b).

Usually the retrospective analysis will be followed by Stage 2 monitoring of future observations, using parameter estimates from the initial group of observations. If assignable causes can be found for out-of-control observations in the initial data, those observations should be removed before estimating the statistical parameters of the process.

Alternative Estimators

Suppose there are m statistically independent observations from the p-variate normal distribution with common covariance matrix,

$$\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}),\ i = 1, \ldots, m.$$

Arranging the observations as row vectors in an $m \times p$ data matrix gives

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_m' \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,p} \end{bmatrix}.$$

Six alternative estimators of the common covariance matrix $\boldsymbol{\Sigma}$ are defined in this section. They are compared in the following section on the basis of the performance of the resulting retrospective Hotelling's T^2 control chart. The first estimator is the familiar sample covariance matrix of the pooled observations

$$\mathbf{S}_1 = \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{x}_i - \overline{\mathbf{x}}) \times (\mathbf{x}_i - \overline{\mathbf{x}})'. \qquad (1)$$

Borrowing from the situation in which there are rational subgroups, the second and third estimators introduce artificial groupings that allow "within" estimates of $\boldsymbol{\Sigma}$. The second estimator is calculated by partitioning the observations into groups of size $p + 1$. Some groups may be larger if necessary so all observations are included in exactly one group. The sample covariance matrix is calculated for each group, and the results are averaged (weighted by the degrees of freedom if the groups differ in size) to give $\mathbf{S}_2$. Since the groups are independent the distribution of the corresponding T^2 statistic is easily calculated. A group size different from $p+1$ could be used. Increasing the size makes the estimate more accurate under the in-control condition, but increases the adverse effect of a shift in the mean vector. Smaller sizes of 2 to p can be used as well, and the average of the group covariance matrices will be full rank when the number of groups is sufficiently large, as shown in the Appendix.

The third estimator $\mathbf{S}_3$ is similar to the second, except overlapping groups are used, which is later shown to increase the effectiveness of the corresponding control chart. Group k in this case contains observations numbered $k, k+1, \ldots, k+r-1$, for $1 \leq k \leq m-r+1$, with r representing the uniform group size. The sample covariance matrices for the groups are statistically independent for estimator two but not for estimator three.

Estimators four and five are variations on the moving range approach used in univariate applications. Estimator four is formed by partitioning the data into independent, non-overlapping groups of two, discarding the last observation if necessary. The difference vectors between the observations in each group are calculated as

$$\mathbf{y}_i = \mathbf{x}_{2i} - \mathbf{x}_{2i-1},\ i = 1, \ldots, \left\lfloor \frac{m}{2} \right\rfloor,$$

where $\lfloor\ \rfloor$ denotes the greatest integer function. One-half the average of the sum of outer products of these differences (the sample covariance matrix with the mean vector known) is an estimate for $\boldsymbol{\Sigma}$,

$$\mathbf{S}_4 = \frac{1}{2} \frac{\mathbf{Y}'\mathbf{Y}}{\lfloor \frac{m}{2} \rfloor},$$

where the matrix $\mathbf{Y}$ consists of row vectors of the differences. Since the groups are independent, the distribution of this estimator is easily established.

Estimator five, suggested by Holmes and Mergen (1993), uses the difference between each successive pair of observations. Let the difference between successive observations be

$$\mathbf{v}_i = \mathbf{x}_{i+1} - \mathbf{x}_i,\ i = 1, \ldots, m-1.$$

The estimator for $\boldsymbol{\Sigma}$ is one-half the sample covariance matrix of these differences,

$$\mathbf{S}_5 = \frac{1}{2} \frac{\mathbf{V}'\mathbf{V}}{(m-1)}. \tag{2}$$

Scholz and Tosch (1994) give a Wishart approximation for the distribution of $\mathbf{S}_5$. Wierda (1994b, p. 47) also describes estimators $\mathbf{S}_4$ and $\mathbf{S}_5$, and notes, "the advantage of the use of the non-overlapping sets over the method of overlapping sets is clear: the distribution of T_i^2 can easily be obtained". However, the use of overlapping groups is later shown to give better performance.

Estimators one through five are shown to be unbiased in the Appendix.

For a standard of comparison the true value of the covariance matrix is included as "estimator" six,

$$\mathbf{S}_6 = \boldsymbol{\Sigma}.$$

Methodology and Results

Let the T^2 statistic for observation i using estimator $\mathbf{S}_j$ be

$$T_{j,i}^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}_j^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}),\ i = 1, \ldots, m \tag{3}$$

where $\bar{\mathbf{x}}$ is the sample mean vector.

When the true parameters are used, the distribution of the quadratic form above is proportional to a χ^2 distribution with p degrees of freedom. In Stage 2 when the true parameters are unknown and $\mathbf{x}_i$ is a future observation, the statistic has Hotelling's T^2 distribution that is proportional to an F distribution,

$$T_{1,i}^2 \frac{m(m-p)}{p(m-1)(m+1)} \sim F(p, m-p). \tag{4}$$

On the other hand, in retrospective Stage 1 the observation $\mathbf{x}_i$ is used in estimating the parameters so the distribution of the statistic is proportional to a beta distribution,

$$T_{1,i}^2 \frac{m}{(m-1)^2} \sim B\left(\frac{p}{2}, \frac{m-p-1}{2}\right). \tag{5}$$

(See Wierda (1994a) or Tracy, Young, and Mason (1992).) Note that Holmes and Mergen (1993) incorrectly apply the control limit derived from an approximation of (4) to their retrospective example and use the same control limit for both of their T^2 charts. Thus, the overall false alarm rate for their conventional T^2 control chart was less than one-seventh of the specified value.

However, in neither Stage 1 or 2 are these statistics independent for different values of i. All share common estimates of the mean vector and covariance matrix, which is especially significant with small sample sizes. Therefore, when highly accurate false alarm probabilities are desired, such as for comparing the performance of alternative control charts, simulation must be used even though the marginal distribution for the statistic may be known. In practice, however, when a single chart is being used, acceptably accurate control limits may be obtained from the marginal distribution. An approximate marginal distribution for the T^2 statistic using $\mathbf{S}_5$ was suggested by Timm (1995),

$$T_{5,i}^2 \frac{m}{(m-1)^2} \sim B\left(\frac{p}{2}, \frac{f-p-1}{2}\right) \tag{6}$$

where

$$f = \frac{2(m-1)^2}{3m-4} \tag{7}$$

as given by Scholz and Tosch (1994).

The control chart using "estimator" six also used the sample mean vector, so simulation was used to determine its control limit.

The performance of the six alternative control charts was compared by simulation with $m = 30$ observations of $p = 2$ dimensional data. Limited comparisons using other values of m and p yielded similar conclusions. Version 2.2 of Mathematica was used for the calculations and plots (Wolfram (1991)).

As shown in the Appendix, the signal probability using any of the alternative estimators of $\mathbf{\Sigma}$ depends on the out-of-control distribution only through the non-centrality parameter

$$(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)'\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0).$$

Therefore, without loss of generality, the identity matrix was used as $\mathbf{\Sigma}$ and $\boldsymbol{\mu}_0$ was the zero vector. The control limits for the six charts were estimated from 3500 simulations for an overall false alarm rate of 0.05 for 30 in-control observations. The control limits used were 10.63, 12.58, 11.91, 14.73, 12.41, 12.49 for the charts corresponding to estimators one through six, respectively.

Step Shift in the Mean Vector

First, consider a step change in the mean vector about midway in the data. A shift after k observations can be modeled as

$$\boldsymbol{\mu}_i = \begin{cases} \boldsymbol{\mu}_0, & i = 1, \ldots, k \\ \boldsymbol{\mu}_0 + \boldsymbol{\delta}, & i = k+1, \ldots, m. \end{cases}$$

In this situation the in-control mean vector is $\boldsymbol{\mu}_0$, and the non-centrality parameter of the observations subject to the special cause is $\boldsymbol{\delta}'\mathbf{\Sigma}^{-1}\boldsymbol{\delta}$.

Estimators two and four are sensitive to the exact location of the change in the mean vector since shifts coincident with a group boundary do not affect the "within" estimate of $\mathbf{\Sigma}$. Therefore, considering only a single location for the change would inaccurately portray the performance of these estimators. Accordingly, the results for step changes after 14, 15, or 16 in-control observations were averaged for estimator two. Estimator four depends on whether the location of the shift is odd or even. Therefore, the performance for the single odd location 15 was counted twice in the averaging for that estimator.

In the simulation 1200 sets of data (giving a standard error less than 0.015) were generated for the following non-centrality parameter values: 5, 10, 15, 20, 25, 30, and 35. Figure 1 shows the estimated signal probability for the six control charts. The usually recommended chart based on $\mathbf{S}_1$ is insensitive regardless of the degree of shift. Thus, this method is completely ineffective in detecting this out-of-control condition *regardless of the severity.* This characteristic is particularly detrimental for high dimensionality data. Even a large shift in the mean vector, as measured by the non-centrality parameter, may not be apparent to someone visually inspecting the data, and statistical analysis may be the only way to detect such shifts.

The two estimators using overlapping groups, $\mathbf{S}_3$ and $\mathbf{S}_5$, provide greater statistical power than the corresponding estimators using independent groups of data, $\mathbf{S}_2$ and $\mathbf{S}_4$. As might be expected, use of the true value of the covariance matrix gives the most powerful T^2 chart.

Having the step change take place near the middle of the observations is the most unfavorable location for the pooled estimator $\mathbf{S}_1$. Its performance would improve in detecting a step change near the beginning or end of the observations. Nevertheless, it seems unwise to use this method because it is insensitive for most shift locations.

Ramp Shift in the Mean Vector (Drift)

Another common model for out-of-control data is a ramp in the mean vector, also called a drift or trend, with the mean vector changing by the same amount for each observation,

$$\boldsymbol{\mu}_i = \boldsymbol{\mu}_0 + \frac{i-1}{m-1}\boldsymbol{\delta}, \; i = 1, \ldots, m,$$

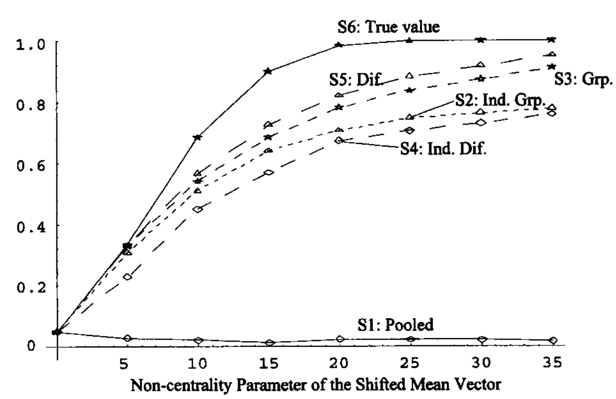

FIGURE 1. Signal Probability for a Step Shift in Mean Vector after 15 of 30 Observations.

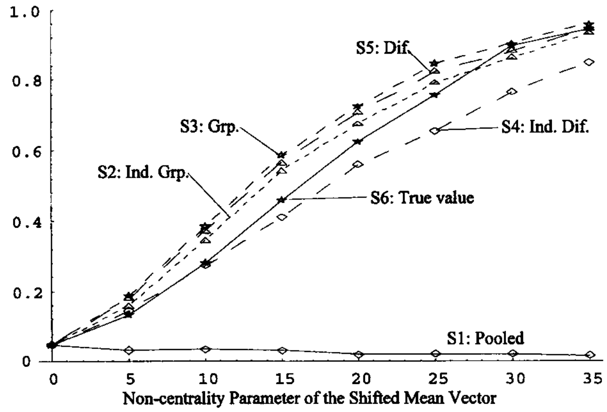

FIGURE 2. Signal Probability for a Ramp Shift in Mean Vector, 30 Observations.

where μ_0 is the in-control mean vector, and δ is the vector difference between the mean vectors of the last and first observations.

Figure 2 shows the estimated signal probability versus the non-centrality parameter of the final observation, again based on simulation with 1200 replications. In this situation also the chart using $\mathbf{S}_1$ fails to show an increase in signal probability as the severity of the shift increases. As before, the estimators using overlapping groups produce greater statistical power than the corresponding estimators using independent groups. Somewhat surprisingly, the charts using estimators $\mathbf{S}_2$, $\mathbf{S}_3$, and $\mathbf{S}_5$ have greater statistical power than that obtained by using the true value of the covariance matrix, $\mathbf{S}_6$. The rationale for this rather odd characteristic, which also holds in the univariate case, has not been determined.

For either the step or ramp shift in the mean vector with a small amount of data, a fairly large shift is required for a large probability of detection with a T^2 chart, similar to the characteristics of a univariate Shewhart chart.

Outliers

Another out-of-control situation consists of k outliers scattered randomly among the m observations. Figure 3 shows the contours of constant out-of-control signal probability for the T^2 charts for zero through five outliers and different severities of the shift in the mean vector affecting only the "outliers". For simplicity all the outliers are presumed here to share a common mean vector. Although strictly speaking the number of outliers must be an integer, these plots were produced by interpolating between the integer values. None of the feasible T^2 charts are effective in detecting outliers. It is worth noting that for this case $\mathbf{S}_1$ and sample covariance matrix of subgroups estimators produce somewhat better results than the difference-based estimators. This is to be expected since *each* isolated outlier will disrupt *two* of the vector differences. Only "estimator" six, the true value, has the desirable property that the signal probability increases with both quantity and severity of outliers. With the others, an increasing number of outliers affects the estimate of $\mathbf{\Sigma}$ so much that the signal probability decreases.

In general, statistical techniques for detecting departures from the in-control condition are known to differ in their effectiveness depending on the nature of the departure. "When ... a specific kind of deviation is feared ... it is appropriate to employ a procedure especially sensitive to that possibility and to use it *in addition* to the overall Shewhart chart (Box and Ramirez (1992))". Such a method, suitable

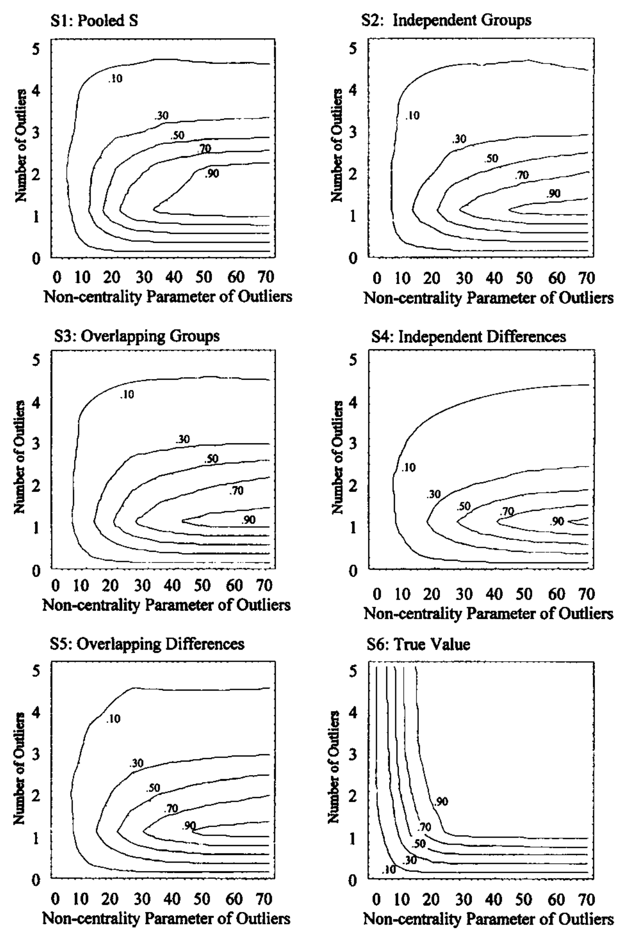

FIGURE 3. Contour Plots of Constant Probability of Out-of-Control Signal with the Indicated Number of Randomly Located Outliers.

for augmenting a T^2 chart for detecting outliers, was proposed by Atkinson and Mulira (1993) and called the *stalactite chart.* This method produces a visual display requiring interpretation. It can be modified as suggested later to provide a numerical threshold for generating an out-of-control signal.

Briefly, the Atkinson and Mulira (1993) multiple-step method consists of selecting a small subset of the data that is hoped to be free of any outliers. The sample mean vector and covariance matrix for the observations in this subset are calculated. Then these statistics are used to select a slightly larger subset of the data that is also hopefully free of outliers. Increasingly larger subsets are selected until the final subset which includes all the data. Note that although the *number* of observations in the subset increases with each step, a *specific* observation in one step may not be retained in the next step. At each step observations with large Mahalanobis distance

$$d_{k,i}^2 = (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)' \, \widehat{\boldsymbol{\Sigma}}_k^{-1} \, (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k) \tag{8}$$

(using the parameters estimated from the current subset with k observations) are flagged as potential outliers. The stalactite chart is an array showing the flagged status of each observation for each step. Outliers tend to be flagged at each step until a step is reached where the subset is so large that some of the outliers must be included. From that point even outliers tend to be unflagged because the covariance matrix is badly estimated. On the other hand, points that are not outliers may be erroneously flagged in the early steps of the process, but tend to remain unflagged in middle and later steps. Normalizing the flagging threshold, as described by Atkinson and Mulira, reduces the extent of false flagging in the early steps and often produces a completely unflagged chart when there are no outliers.

Atkinson and Mulira recommend starting the process with a randomly selected subset of $p+1$ observations, although there is the possibility of including outliers. Several randomly selected starting subsets may be used to increase the chance of starting with an outlier-free subset. Atkinson and Mulira report this does not seem to be necessary since an observation in a subset can be dropped from later subsets. They report the algorithm usually recovers even when outliers are present in the initial subset. They give an example in which the initial subset includes only outliers, and the algorithm still performs properly.

At each step of the algorithm the "hopefully outlier-free" subset is determined as follows. The initial subset consists of $p+1$ randomly selected observations. Consider the general case for step k in which there are k observations in the subset. The sample mean vector and covariance matrix for this subset are used to calculate the Mahalanobis distances for all m observations. The subset for the next step is formed by taking the $k+1$ observations having the smallest Mahalanobis distances and so on until all m observations are included in the final subset. In the unlikely event that outliers dominate the initial subset, a step will eventually be reached in which the in-control observations dominate. This will tend to exclude all the outliers from subsequent subsets. Of course, when the number of outliers is overwhelming, this procedure may flag the in-control points instead of the outliers. Figure 4 shows stalactite charts for 30 observations with zero, five, and ten outliers, all of which are easily recognizable.

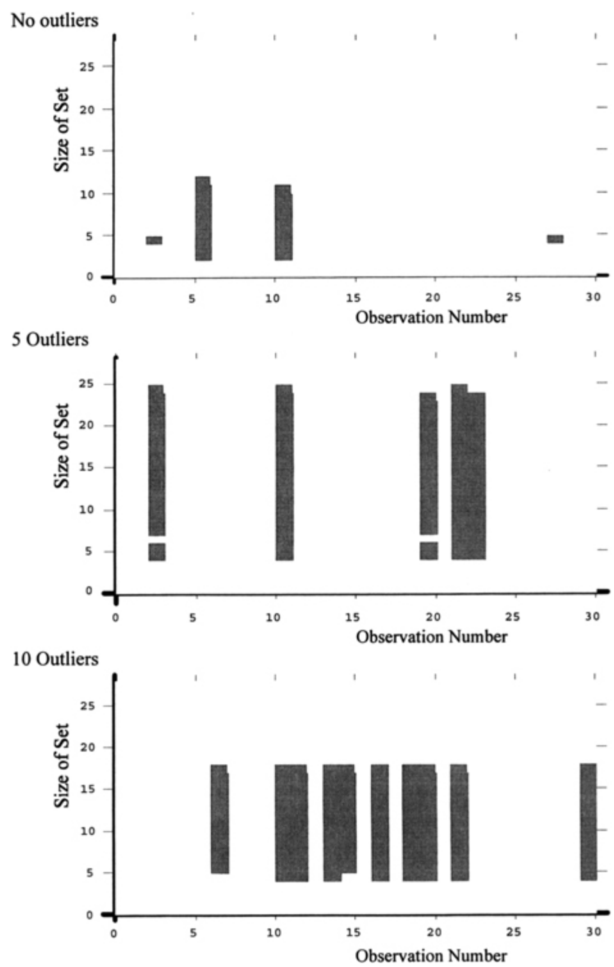

FIGURE 4. Representative Stalactite Plots with 0, 5, and 10 Outliers.

Although the stalactite chart is powerful for visually detecting a pattern consistent with outliers, interpretation skill is required. Outliers tend to be unflagged in the later stages of the algorithm, and the last step depends on the number of outliers present. Thus, the observer needs to form a subjective assessment about the number of outliers and the step at which the outliers begin to be unflagged. Since this process does not lend itself to computerized decision making, it is difficult to evaluate the statistical power of the technique as proposed by Atkinson and Mulira.

Accordingly, the Atkinson and Mulira method was modified by incorporating a computerized decision rule. In a data set of 30 observations, Figure 3 shows the feasible T^2 charts are not effective with as many as *three* outliers. Therefore, a procedure reliably detecting *five* outliers in 30 observations would be a considerable improvement. We tailored the method to be sensitive to five or fewer outliers, at the expense of sensitivity to a greater number of outliers. Specifically, a single statistic was considered, the maximum Mahalanobis distance for the step in which all but five observations are included in the subset. When this statistic was larger than an upper control limit (determined by simulation) an out-of-control signal was generated. Figure 5 shows the contours of constant signal probability for this scheme. We estimated the signal probability with 400 simulations for each combination of number of outliers and noncentrality parameter that are shown on the axes of the figure. It turns out the resulting chart is effective with even more than five outliers, reliably detecting as many as seven outliers in a group of 30 with large shifts. Even more outliers could be detected, if desired, by using the statistic from an earlier step of the algorithm, or by using the visual method of Atkinson and Mulira. By comparing Figure 5 to Figure 3, it is clear that the stalactite method is much more effective in detecting outliers than any of the feasible T^2 charts.

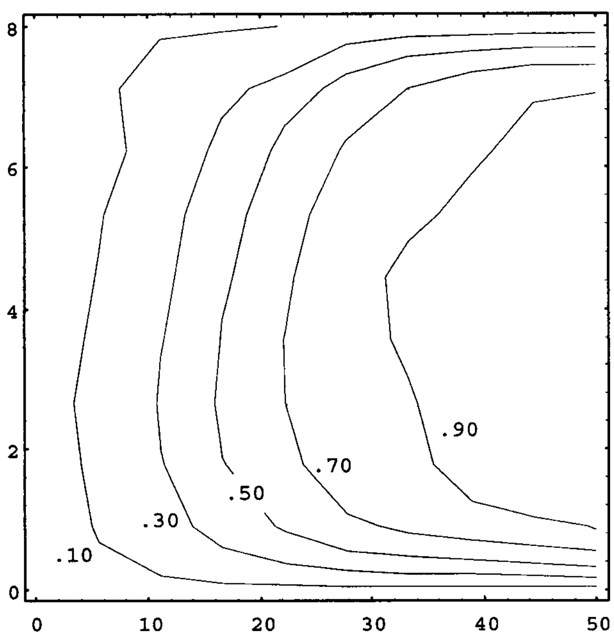

FIGURE 5. Contour Plot of Constant Probability of an Out-of-Control Signal with Randomly Located Outliers using a Modification of the Stalactite Method.

With a large step shift in the mean vector, taking overlapping differences is likely to produce a set of vectors with a single outlier, so detecting this outlier could improve the performance in detecting a step shift. As noted above, the sample covariance matrix of these differences is fairly robust for a single outlier. However, viewed in this light, other techniques are possible. A common intuitive idea for coping with a single outlier is the backward elimination method of considering all subsets with a single observation deleted. (For example, see Roes, Does, and Schurink (1993).) Atkinson and Mulira (1993) show that using this "deletion Mahalanobis distance",

$$\left(\mathbf{x}_i - \bar{\mathbf{x}}^*\right)' \left(\mathbf{S}^*\right)^{-1} \left(\mathbf{x}_i - \bar{\mathbf{x}}^*\right),$$

where the mean vector and covariance matrix are estimated with one observation deleted, "is a monotone function of the usual Mahalanobis distance and so provides no additional diagnostic information". This is in contrast to the regression diagnostic situation in which the deletion distance is helpful in identifying single outliers. Therefore, the regular Mahalanobis distance can be used to identify the most suspect of the differences, the one having the largest distance. Alternatively, the Atkinson and Mulira stalactite method could be employed. Regardless of the method used to detect the possible outlier, deleting the difference vector most likely to have been affected by a step shift and using only the remaining difference vectors to estimate the covariance matrix should be fairly robust against a step shift in the mean vector. On the other hand, unless the step shift is large, the presence of a single outlier will have little effect on the sample covariance matrix, and the technique described above is already effective in detecting large step shifts. Therefore, the improvement in sensitivity may not be worth the additional computation.

Example

Table 1 reproduces the data analyzed by Holmes and Mergen (1993) which gives the composition of "grit" manufactured by a plant in Europe. The columns headed L, M, and S give the percentage classified as large, medium, and small, respectively, for each of 56 observations. Since a dependency exists, only the first two components of each observation were used. For this data the mean vector is $[5.682, 88.22]'$. The estimators one and five have the following values:

Sample Covariance Matrix, $\mathbf{S}_1$

$$\begin{bmatrix} 3.770 & -5.495 \\ -5.495 & 13.53 \end{bmatrix}$$

Sample Covariance of Successive Differences, $\mathbf{S}_5$

$$\begin{bmatrix} 1.562 & -2.093 \\ -2.093 & 6.721 \end{bmatrix}$$

The T^2 statistics using these estimates appear in Table 1 and the corresponding control charts are given in Figures 6 and 7. The control limits, determined from 2000 simulations are 10.55 and 11.35, respectively, for the charts using estimators one and five. These control limits correspond to a 0.155 overall false alarm probability, which corresponds to that of a Shewhart chart with 56 independent observations, each with a false alarm probability of 0.003.

Only the control chart in Figure 7 signals. A stalactite analysis for outliers, not shown, was negative. In classifying the out-of-control pattern represented by these data, we note that the mean vectors of the first 24 and final 32 observations are, respectively, $[4.23, 90.8]'$ and $[6.77, 86.3]'$, which are statistically significantly different, while the "within" covariance matrices are not statistically different. This is evidence of a shift in the mean vector following observation 24. Holmes and Mergen (1993) attributed the out-of-control signal to outliers, specifically observations 26 and 52, the ones for which the T^2 statistic exceeded the control limit. Although it is true that outliers would tend to show up in this manner, they are more likely to appear on the chart using $\mathbf{S}_1$, since

TABLE 1. Example from Holmes and Mergen (1993) showing the Data and the T^2 Statistics using Estimators $\mathbf{S}_1$ and $\mathbf{S}_5$

i	$L = x_{i,1}$	$M = x_{i,2}$	$S = x_{i,3}$	$T^2_{1,i}$	$T^2_{5,i}$	i	$L = x_{i,1}$	$M = x_{i,2}$	$S = x_{i,3}$	$T^2_{1,i}$	$T^2_{5,i}$
1	5.4	93.6	1.0	4.496	6.439	29	7.4	83.6	9.0	1.594	3.261
2	3.2	92.6	4.2	1.739	4.227	30	6.8	84.8	8.4	0.912	1.743
3	5.2	91.7	3.1	1.460	2.200	31	6.3	87.1	6.6	0.110	0.266
4	3.5	86.9	9.6	4.933	7.643	32	6.1	87.2	6.7	0.077	0.166
5	2.9	90.4	6.7	2.690	5.565	33	6.6	87.3	6.1	0.255	0.564
6	4.6	92.1	3.3	1.272	2.258	34	6.2	84.8	9.0	1.358	2.069
7	4.4	91.5	4.1	0.797	1.676	35	6.5	87.4	6.1	0.203	0.448
8	5.0	90.3	4.7	0.337	0.645	36	6.0	86.8	7.2	0.193	0.317
9	8.4	85.1	6.5	2.088	4.797	37	4.8	88.8	6.4	0.297	0.590
10	4.2	89.7	6.1	0.666	1.471	38	4.9	89.8	5.3	0.197	0.464
11	3.8	92.5	3.7	1.368	3.057	39	5.8	86.9	7.3	0.242	0.353
12	4.3	91.8	3.9	0.951	1.986	40	7.2	83.8	9.0	1.494	2.928
13	3.7	91.7	4.6	1.105	2.688	41	5.6	89.2	5.2	0.136	0.198
14	3.8	90.3	5.9	1.019	2.317	42	6.9	84.5	8.6	1.079	2.062
15	2.6	94.5	2.9	3.099	7.262	43	7.4	84.4	8.2	1.096	2.477
16	2.7	94.5	2.8	3.036	7.025	44	8.9	84.3	6.8	2.854	6.666
17	7.9	88.7	3.4	3.803	6.189	45	10.9	82.2	6.9	7.677	17.666
18	6.6	84.6	8.8	1.167	1.997	46	8.2	89.8	2.0	6.677	10.321
19	4.0	90.7	5.3	0.751	1.824	47	6.7	90.4	2.9	2.708	3.869
20	2.5	90.2	7.3	3.966	7.811	48	5.9	90.1	4.0	0.888	1.235
21	3.8	92.7	3.5	1.486	3.247	49	8.7	83.6	7.7	2.424	5.914
22	2.8	91.5	5.7	2.357	5.403	50	6.4	88.0	5.6	0.261	0.470
23	2.9	91.8	5.3	2.094	4.959	51	8.4	84.7	6.9	1.995	4.731
24	3.3	90.6	6.1	1.721	3.800	52	9.6	80.6	9.8	4.732	11.259
25	7.2	87.3	5.5	0.914	1.791	53	5.1	93.0	1.9	2.891	4.303
26	7.3	79.0	13.7	9.226	14.372	54	5.0	91.4	3.6	0.989	1.609
27	7.0	82.6	10.4	2.940	4.904	55	5.0	86.2	8.8	1.770	2.495
28	6.0	83.5	10.5	3.310	4.771	56	5.9	87.2	6.9	0.102	0.166

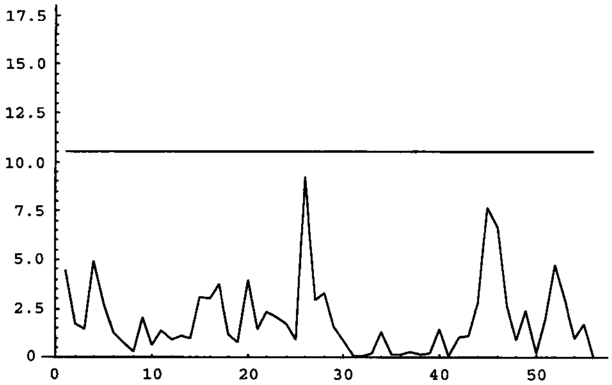

FIGURE 6. T^2 Control Chart using Estimator $\mathbf{S}_1$.

it is more sensitive to outliers. Furthermore, a shift in the mean vector would produce a similar pattern, with only one or a few observations exceeding the control limit.

Rational Subgroups Larger than One

There may be inherent process characteristics that cause homogeneity over a group of several observations, a rational subgroup. For example, samples from the same batch of a chemical process are apt to comprise a rational subgroup. When there is reason to believe there are subgroups having a homogeneous distribution, it is standard practice to average the sample covariance matrices for each subgroup to estimate the common covariance matrix. (For example, see Montgomery (1996), Alt and Smith (1988), or Fuchs and Benjamini (1994).) The subgroups may not all be the same size; observations may be missing from some of the subgroups. When there are different numbers of observations in the subgroups, the average of their sample covariance matrices should be calculated using their degrees of freedom as weights.

Sometimes a distinction is unnecessarily made according to whether the subgroup size is greater than p or not. Since the sample covariance matrix for less than $p + 1$ observations is singular, the recommendation is sometimes made that when the subgroup size n (assumed to be the same for all subgroups) is less than $p + 1$, the observations are to be treated as individuals. The sample covariance matrix of the pooled observations is used to estimate $\mathbf{\Sigma}$. (See Wierda (1994a), corrected in Wierda (1994b, pp. 29–30).) This distinction according to the subgroup size is unnecessary. Even if some or all the subgroup sample covariance matrices are less than full rank, the degrees of freedom of the independent sample covariance matrices are additive. Therefore, as shown in the Appendix, the average of the subgroup sample covariance matrices will be full rank provided $p < m(n - 1)$.

Nevertheless, there is seldom an ironclad guarantee that all the observations in a rational subgroup have the same mean vector. Therefore, even with rational subgroups, it may be prudent to consider the possibility of a shift in the mean vector *within* a rational subgroup, especially if the subgroups are large. For this situation, it would be appropriate to use the overlapping differences *within* each subgroup, with the time order of the observations in the subgroup maintained. Estimate five, given by (2), would be calculated for each subgroup. Then the average of these estimates (weighted by degrees of freedom) would be used to estimate $\mathbf{\Sigma}$. This has the advantage that shifts in the mean vector outside a subgroup would not upset the estimation, and the effect of a step change inside one of the subgroups would be reduced compared to using the sample covariance of the observations themselves. On the other hand, this approach would be less sensitive to outliers occurring within the (presumably) rational subgroups.

Conclusions

In the retrospective Stage 1 situation special care is in order because the analyst is simultaneously deciding if the process is in statistical control and estimating the in-control parameters of the process. When the process does not provide a reasonable basis for rational subgroups, the standard recommendation for multivariate analysis uses the sample covariance matrix to estimate the covariance matrix. This estimate and the sample mean vector are used in a Hotelling's T^2 control chart. We have shown that this recommendation is not effective in detecting a step or ramp shift in the mean vector, even with a very large shift. Although the results are depicted

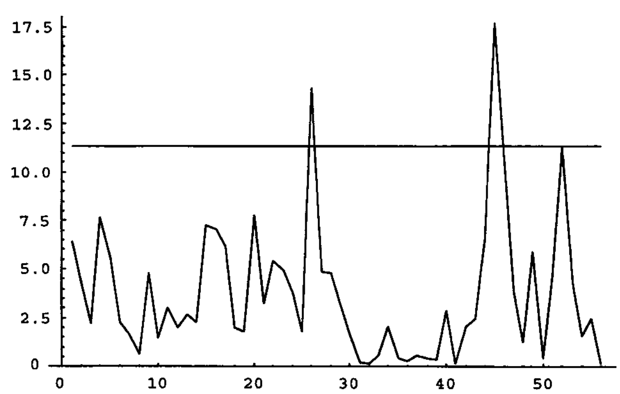

FIGURE 7. T^2 Control Chart using Estimator $\mathbf{S}_5$.

for a simulation with a specific number of observations and dimensionality, there is no reason to believe the results do not apply to the general case. Other simulations not reported here support this conclusion.

A more effective chart estimates the covariance matrix from the vector differences between successive observations, as suggested by Holmes and Mergen (1993), under the presumption that successive observations will tend to have the same, or nearly the same, mean vector. The sample covariance matrix for these vector differences divided by two, as given in (2), is a more robust estimate of the in-control covariance matrix. The corresponding T^2 control chart is effective in detecting shifts in the mean vector.

Although the procedure using (2) to estimate the sample covariance matrix is effective in detecting a step or ramp shift in the mean vector, it is not effective in detecting outliers. When outliers may be present, the method suggested by Atkinson and Mulira (1993) or the modification suggested herein incorporating a computerized decision rule should be used. This method is swamped by a step shift in the mean vector near the middle of the observations, so the overlapping differences method should be used as well to detect step or ramp shifts.

Appendix

Average of Rank Deficient Sample Covariance Matrices is Full Rank

When there are less than $p+1$ observations in a group, the sample covariance matrix will be less than full rank. Nevertheless, the average for all the groups will be full rank (unless there is very little data). To establish this fact, suppose all groups are the same size and there are n observations in each group. The sample covariance matrix for each group times $n-1$ will be distributed as a Wishart matrix with $n-1$ degrees of freedom (and singular if $n < p+1$). However, the sum of these Wishart matrices with a common scale matrix is also Wishart. The degrees of freedom of the sum will be $m(n-1)$, the sum of the degrees of freedom. Provided the total degrees of freedom exceeds p, the weighted average of the sample covariances will be non-singular with probability one (Mardia, Kent, and Bibby (1979, Theorem 3.4.4 and Theorem 3.4.8)). (Note that Corollary 3.4.8.1 gives the sample covariance matrix as non-singular when p is *less than* the degrees of freedom, whereas it should read *less than or equal to*.)

The Successive Differences Estimator is Unbiased

Estimators four and five are unbiased for the in-control situation where the observations are identically distributed. To establish this fact, let matrix $\mathbf{E}$ be the $m-1 \times m$ matrix

$$\mathbf{E} = \begin{bmatrix} -1 & 1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -1 & 1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -1 & 1 \end{bmatrix}$$

Also, let the vectorized form of $\mathbf{X}'$ be the $mp \times 1$ column vector formed by stacking the columns of the matrix $\mathbf{X}'$,

$$(\mathbf{X}')^{\text{Vec}} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_m \end{bmatrix}.$$

The vectorized form of the transpose of the matrix $\mathbf{V}$ can be alternatively defined as

$$\begin{aligned} (\mathbf{V}')^{\text{Vec}} &= \begin{bmatrix} -\mathbf{I}_p & \mathbf{I}_p & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_p & \mathbf{I}_p & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{I}_p & \mathbf{I}_p \end{bmatrix} (\mathbf{X}')^{\text{Vec}} \\ &= (\mathbf{E} \otimes \mathbf{I}_p) \cdot (\mathbf{X}')^{\text{Vec}} \end{aligned}$$

using the Kronecker product notation. Equivalently,

$$\begin{aligned} (\mathbf{V}') &= (\mathbf{I}_p \cdot \mathbf{X}' \cdot \mathbf{E}')^{\text{Vec}} \\ \mathbf{V} &= \mathbf{E} \cdot \mathbf{X}. \end{aligned}$$

The covariance matrix for $(\mathbf{X}')^{\text{Vec}}$ is $(\mathbf{I}_m \otimes \boldsymbol{\Sigma})$, so the covariance matrix for $(\mathbf{V}')^{\text{Vec}}$ is

$$(\mathbf{E} \otimes \mathbf{I}_p)(\mathbf{I}_m \otimes \boldsymbol{\Sigma})(\mathbf{E}' \otimes \mathbf{I}_p) = (\mathbf{E} \cdot \mathbf{E}') \otimes \boldsymbol{\Sigma}.$$

Thus, the expected value of the outer product of a row of $\mathbf{V}$ with itself is $2\boldsymbol{\Sigma}$, for any of the $m-1$ rows. Therefore, the expected value of $\mathbf{V}'\mathbf{V}$ is $2(m-1)\boldsymbol{\Sigma}$. A similar method shows estimator four to be unbiased as well. Furthermore, the sample covariance matrices are unbiased for the in-control situation, so estimators one, two, and three are also unbiased if the process is in-control.

Invariance of T^2 Statistics to Linear Transformations

When using Hotelling's T^2 statistic, regardless of the estimator used for $\boldsymbol{\Sigma}$, the quadratic form is invariant to a linear transformation of the data. Estimators one, two, and three use the sample covariance matrix, so the invariance of Hotelling's T^2 statistic with these estimators is established. To demonstrate the invariance for estimator five, which is defined in (2), consider a linear transformation

$$\mathbf{z}_i = \mathbf{\Lambda x}_i + \boldsymbol{\lambda}$$
$$\mathbf{Z} = \mathbf{X\Lambda}' + \mathbf{1}_m\boldsymbol{\lambda}'$$

where $\mathbf{z}_i \sim N_p(\boldsymbol{\mu} + \boldsymbol{\lambda}, \mathbf{\Lambda\Sigma\Lambda}')$.

Calculating Hotelling's T^2 statistic for transformed observation $\mathbf{z}_i$ (assuming $\mathbf{\Lambda}$ is full rank) gives

$$\begin{aligned} q_i^2 &= (\mathbf{z}_i - \bar{\mathbf{z}})' \left(\frac{\mathbf{Z'E'EZ}}{2(m-1)}\right)^{-1} (\mathbf{z}_i - \bar{\mathbf{z}}) \\ &= (\mathbf{\Lambda x}_i + \boldsymbol{\lambda} - \mathbf{\Lambda}\,\bar{\mathbf{x}} - \boldsymbol{\lambda})' \left(\frac{\mathbf{\Lambda X'E'EX\Lambda}'}{2(m-1)}\right)^{-1} \\ &\quad \times (\mathbf{\Lambda x}_i + \boldsymbol{\lambda} - \mathbf{\Lambda}\,\bar{\mathbf{x}} - \boldsymbol{\lambda}) \\ &= (\mathbf{x}_i - \bar{\mathbf{x}})' \left(\frac{\mathbf{X'E'EX}}{2(m-1)}\right)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}). \end{aligned}$$

Thus, the T^2 statistic for this observation takes the same value regardless of whether the original or the transformed variable is used, although the mean vector and estimate of the covariance matrix might be very different. Similar methods show the quadratic form is also invariant when estimate four or six is used.

Acknowledgments

We would like to thank Dr. Bruce Barrett of the University of Alabama for his help with the problem of detecting outliers, Dr. Neil Timm of the University of Pittsburgh for suggesting the approximate marginal distribution for T_5^2, Dr. Siebrand Wierda of the University of Groningen for helpful comments on an earlier version of this paper, and the referees for several significant improvements in the content and presentation.

References

ALT, F. B. and SMITH, N. D. (1988). "Multivariate Process Control" in *Handbook of Statistics, Vol. 7* edited by P. R. Krishnaiah and C. R. Rao. North-Holland, Amsterdam. pp. 333–351.

ATKINSON, A. C. and MULIRA, H. M. (1993). "The Stalactite Plot for the Detection of Multivariate Outliers". *Statistics and Computing* 3, pp. 27–35.

BOX, G. E. P. and RAMIREZ, J. (1992). "Cumulative Score Charts". *Quality and Reliability Engineering International* 8, pp. 17–27.

FUCHS, C. and BENJAMINI, Y. (1994). "Multivariate Profile Charts for Statistical Process Control". *Technometrics* 36, pp. 182–195.

HOLMES, D. S. and MERGEN, A. E. (1993). "Improving the Performance of The T^2 Control Chart". *Quality Engineering* 5, pp. 619–625.

JACKSON, J. E. (1985). "Multivariate Quality Control". *Communications in Statistics: Theory and Methods* 11, pp. 2657–2688.

LOWRY, C. A. and MONTGOMERY, D. C. (1995). "A Review of Multivariate Control Charts". *IIE Transactions* 27, pp. 800–810.

MARDIA, K. V.; KENT, J. T.; and BIBBY, J. M. (1979). *Multivariate Analysis.* Academic Press, London.

MONTGOMERY, D. C. (1996). *Introduction to Statistical Quality Control,* 3rd ed. John Wiley & Sons, New York, NY.

NELSON, L. S. (1982). "Control Charts for Individual Measurements". *Journal of Quality Technology* 14, pp. 172–173.

ROES, K. C. B.; DOES, R. J. M. M.; and SCHURINK, Y. (1993). "Shewhart-Type Control Charts for Individual Observations". *Journal of Quality Technology* 25, pp. 188–198.

SCHOLZ, F. W. and TOSCH, T. J. (1994). "Small Sample Uni- and Multivariate Control Charts for Means". *Proceedings of the American Statistical Association, Quality and Productivity Section.*

TIMM, N. H. (1995). Private communication.

TRACY, N. D.; YOUNG, J. C.; and MASON, R. L. (1992). "Multivariate Control Charts for Individual Observations". *Journal of Quality Technology* 24, pp. 88–95.

WIERDA, S. J. (1994a). "Multivariate Statistical Process Control — Recent Results and Directions for Future Research". *Statistica Neerlandica* 48, pp. 147–168.

WIERDA, S. J. (1994b). *Multivariate Statistical Process Control.* Groningen Theses in Economics, Management and Organization. Wolters-Noordhoff, Groningen, The Netherlands.

WOLFRAM, S. (1991). *Mathematica: a System for Doing Mathematics by Computer,* 2nd ed. Addison-Wesley, Redwood City, CA.

Key Words: *Hotelling's T^2 Statistic, Multivariate Control Charts, Outliers, Statistical Process Control.*